Scaling laws of strategic behaviour and size heterogeneity in agent dynamics

Gabriella Vaglica,[1] Fabrizio Lillo,[1,2] Esteban Moro,[3] and Rosario N. Mantegna[1]

[1]*Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, Viale delle Scienze, I-90128, Palermo, Italy*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*
[3]*Grupo Interdisciplinar de Sistemas Complejos (GISC)*
Departamento de Matemáticas, Universidad Carlos III de Madrid,
Avedida de la Universidad 30, E-28911, Leganés, Spain

(Dated: December 5, 2008)

The dynamics of many socioeconomic systems is determined by the decision making process of agents. The decision process depends on agent's characteristics, such as preferences, risk aversion, behavioral biases, etc. [1, 2]. In addition, in some systems the size of agents can be highly heterogeneous leading to very different impacts of agents on the system dynamics [3, 4, 5, 6, 7, 8]. The large size of some agents poses challenging problems to agents who want to control their impact, either by forcing the system in a given direction or by hiding their intentionality. Here we consider the financial market as a model system, and we study empirically how agents strategically adjust the properties of large orders in order to meet their preference and minimize their impact. We quantify this strategic behavior by detecting scaling relations of allometric nature [9] between the variables characterizing the trading activity of different institutions. We observe power law distributions in the investment time horizon, in the number of transactions needed to execute a large order and in the traded value exchanged by large institutions and we show that heterogeneity of agents is a key ingredient for the emergence of some aggregate properties characterizing this complex system.

PACS numbers:

In many complex systems agents self organize themselves in an ecology of different "species" interacting in a variety of ways. Agents are not only different in their strategies, information, and preferences, but they can be very different in their size. Examples include individual's wealth [3] and firms size [5, 6]. The presence of agents with large size poses several challenging questions. It is likely that large agents impacts the system in a way that is significantly different from small ones. Indeed, small agents can easily hide their intentionality, while for large agents this is not so easy and they must adopt strategies taking into account their own effect because revealing their intention could decrease their fitness.

Financial markets are an ideal system to investigate this problem. There is empirical evidence that market participants are very heterogeneous in size. For example banks [7] and mutual funds [8] size follow Zipf's law, i.e. the probability that the size of a participant is larger than x decays as $1/x$ [4]. As a consequence large investors usually need to trade large quantities that can significantly affect prices. The associated cost is called market impact [10, 11, 12, 13, 14, 15]. For this reason large investors refrain from revealing their demand or supply and they typically trade their large orders incrementally over an extended period of time. These large orders are called *packages* [16, 17] or *hidden orders* and are split in smaller trades as the result of a complex optimization procedure which takes into account the investor's preference, risk aversion, investment horizon, etc..

Here we investigate the trading activity of a large fraction of the financial firms exchanging a financial asset at the Spanish Stock Market (Bolsas y Mercados Españoles, BME) in the period 2001-2004 (see Materials and Methods section for a description of data). Firms are credit entities and investment firms which are members of the stock exchange and are entitled to trade in the market.

Our approach aims to be a comprehensive approach analysing the overall dynamics of all packages exchanged in the market. However, our database does not contain direct information on packages, so that this information must be statistically inferred from the available data. Since we do not have information on clients but only on firms, we develop a detection algorithm (see Material and Methods for a description of the algorithm) which is not sensible to small fluctuation in the buy/sell activity of a firm. The algorithm detects time segments in the inventory time evolution of a firm when the firm acts as a net buyer or seller at an approximately constant rate. We call these segments *patches* and we assume that in each of these patches it is contained at least one package.

Since firms act simultaneously as brokers for many clients, it is rather frequent that in a patch not all the transactions have the same sign. However, a vast majority of firm inventory time series can be partitioned in patches with a well defined direction to buy or to sell. This is probably due to the fact that in most cases the trading activity of a firm is dominated by the activity of one big client. We consider *directional patches*, i.e. patches with a well defined direction (see Figure 1). The characterizing variables of a directional patch are the time length T (in seconds) of the patch, measured as the time interval between the first and the last order of the patch, the traded value V_m and the number N_m of trades characterizing the patch. For example, N_m is the number of buy trades and V_m is the purchased value

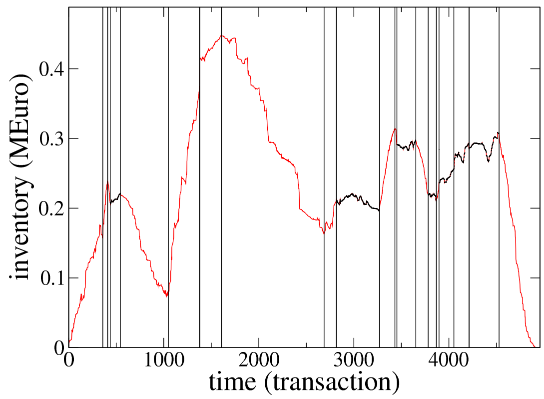

FIG. 1: Example of an inventory time series. The series refers to a particular firm trading Santander. The vertical lines indicate the position where our algorithm predicts the boundary between two patches. The red patches are directional patches. Due to their statistical nature, in each patch there are buy (with a total traded value V_b) and sell (with a total traded value V_s) trades. We consider directional patches, i.e. patches where either $V_b/V > \theta$ (buy patch) or $V_s/V > \theta$ (sell patch), where $V = V_b + V_s$. For buy patches $V_m = V_b$ whereas for sell patches $V_m = V_s$. In the present study we set $\theta = 75\%$. Directional patches are shown as red lines. The black patches are not directional and are not considered in the rest of the paper.

for buy patches.

We investigate first the distributional properties of the patches identified by our algorithm. Figure 2 shows the distribution of T, N_m, and V_m for the three investigated stocks. The asymptotic behavior of all the three distributions can be approximated by a power law function $P(X) \sim 1/X^{\zeta_X+1}$, where X can be T, N_m, or V_m and ζ_X is the exponent characterizing the power law behavior. A summary of the estimated exponents is shown in Table I from which one can conclude that $\zeta_{V_m} \simeq 2$, $\zeta_{N_m} \simeq 1.8$, and $\zeta_T \simeq 1.3$. Our analysis makes explicit the presence of very broad distribution for the three variables characterizing a patch. In fact the very low value of the exponents is consistent with the conclusion that T and N_m belong to the domain of Lévy stable distributions. This result indicates that in the market there is a huge heterogeneity in the scales characterizing the trading profile of the investors.

The volume of the packages is likely to be related to the size of the investors. Large investors need to trade large packages to rebalance their portfolio. Gabaix *et al.* [18] developed a theory which predicts that package size should be power law distributed with an exponent $\zeta_{V_m} = 3/2$. The value we find for $\zeta_{V_m} \simeq 2.$ is slightly larger than the one predicted by them. On the contrary, the value $\zeta_{N_m} = 3$ derived by the theory in [18] is significantly larger than our estimate ($\zeta_{N_m} \simeq 1.8$). Finally, the power law distribution of packages time length T might reflect the heterogeneity of time scales among investors. The distribution of T is compatible with the ones obtained by using specialized database describing the investment packages of large investors [16, 17] (see Figure 2). Gabaix *et al.* theory [18] predicts the value $\zeta_T = 3$ which is significantly larger than our value ($\zeta_T \simeq 1.3$). The presence of power law distribution of investors time scales has been recently suggested in stylized models of investment decisions [19, 20, 21].

The role of size heterogeneity in the emergence of power law distributions will be considered at the end of the paper. To complete our characterization of firm patches, we now consider the relation between the variables characterizing each patch. Specifically, by applying the Principal Component Analysis (PCA) to the set of points with coordinates $(\log T, \log N_m, \log V_m)$, we investigate the allometric relations between any two of the above variables, i.e.

$$N_m \sim V_m^{g_1} \qquad T \sim V_m^{g_2} \qquad N_m \sim T^{g_3} \tag{1}$$

Figure 3 shows the scatter plots and the contour plots for the stock Telefónica. In all three cases a clear dependence between the variables is seen. PCA analysis shows that the first eigenvalue explains on average 91%, 83%, and 89% of the variance for the first, second, and third allometric relation, respectively, indicating a strong correlation between the variables. The estimated exponents (see Table I) are consistent for different stocks so that the allometric relations are

$$N_m \sim V_m^{1.1} \qquad T \sim V_m^{1.9} \qquad N_m \sim T^{0.66} \tag{2}$$

The presence of scaling relations between the variables were first suggested in Ref. [18] but it is worth noting that the theory developed in that paper predicts $g_1 = g_2 = 1/2$ and $g_3 = 1$, and these values are quite different from the ones

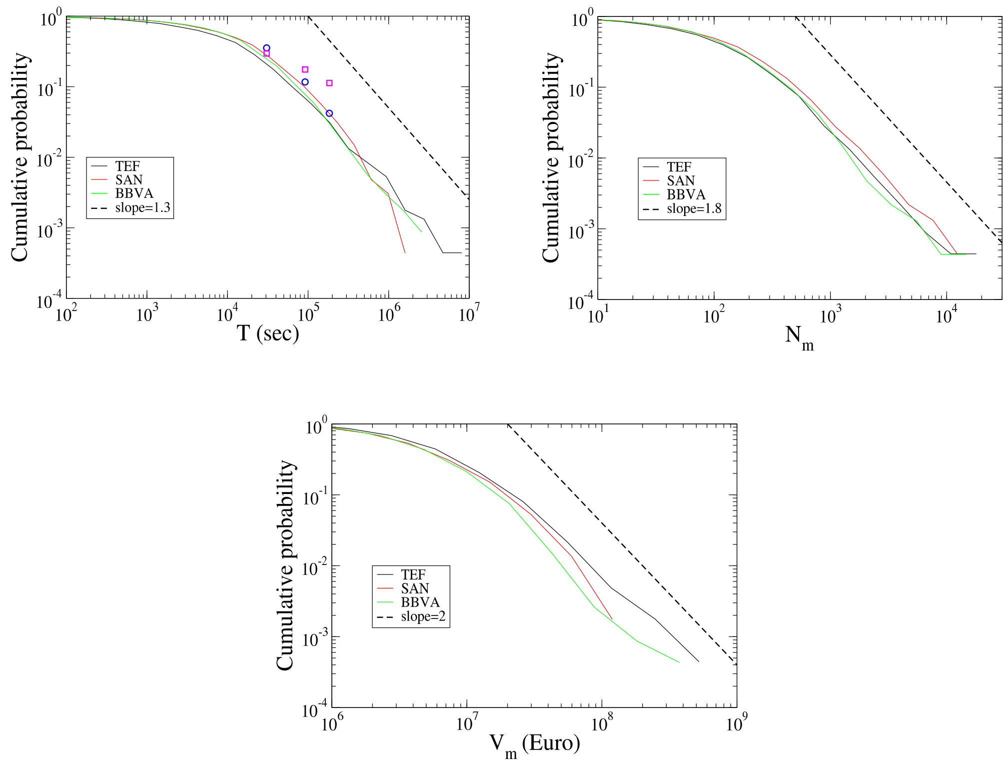

FIG. 2: Distribution of T, N_m, and V_m for the the stocks Banco Bilbao Vizcaya Argentaria (BBVA), Banco Santander Central Hispano (SAN), and Telefónica (TEF). In the panel showing the distribution of T we plot the distribution of packages reported in literature on packages. Specifically, blue circles are results from Ref. [16] for packages traded at the New York Stock Exchange and magenta squares are results from Ref. [17] for packages traded at the Australian Stock Exchange.

TABLE I: Summary of the properties of detected patches. The number in parenthesis nearby the tick symbol is the number of patches detected for the considered stock. Rows 1-3: Tail exponents of the distribution of T, N_m, and V_m estimated with the Hill estimator (or Maximum Likelihood Estimator). In parenthesis we report the 95% confidence interval. Rows 4-6: Exponents of the allometric relations defined in Eq. 1. The exponents are estimated with PCA and the errors are estimated with bootstrap. In parenthesis we report the 95% confidence interval. Rows 7-9: Percentage of firms with at least 10 patches for which one cannot reject the hypothesis of lognormality with 95% confidence according to Jarque-Bera test. The numbers in parenthesis are the number of firms for which one cannot reject the hypothesis of lognormality divided to the number of firms used in the test.

	BBVA (2104)	SAN (2086)	TEF (2062)
ζ_{V_m}	2.3 (1.9 ; 2.7)	2.0 (1.7 ; 2.3)	1.9 (1.6 ; 2.2)
ζ_{N_m}	2.0 (1.7 ; 2.3)	1.7 (1.4 ; 2.0)	1.7 (1.4 ; 2.0)
ζ_T	1.5 (1.3 ; 1.7)	1.5 (1.3 ; 1.7)	1.2 (1.0 ; 1.4)
g_1	1.08 (1.05 ; 1.12)	1.06 (1.01 ; 1.10)	1.07 (1.04 ; 1.11)
g_2	1.81 (1.69 ; 1.93)	1.81 (1.68 ; 1.94)	2.00 (1.88 ; 2.14)
g_3	0.68 (0.65 ; 0.71)	0.68 (0.65 ; 0.70)	0.62 (0.59 ; 0.64)
T	75 (15/20)	63 (17/27)	77 (24/31)
N_m	90 (18/20)	100 (27/27)	100 (31/31)
V_m	90 (18/20)	100 (27/27)	94 (29/31)

we estimate from data. The first allometric relation indicates that the number of transactions in which a package is split is approximately proportional to the total traded value of the package. This implies that the mean transaction volume is roughly independent on the size of the package. This mean value is on average determined by the size of the available volume at the best quote indicating that the trader does not trade orders larger than the volume available at the best quote, probably to avoid being too aggressive [22].

We consider the relation between the three variables together by performing a PCA on the set of points describing the patches and identified by the coordinates $(\log T, \log N_m, \log V_m)$ [23]. The set of points effectively lies on a two

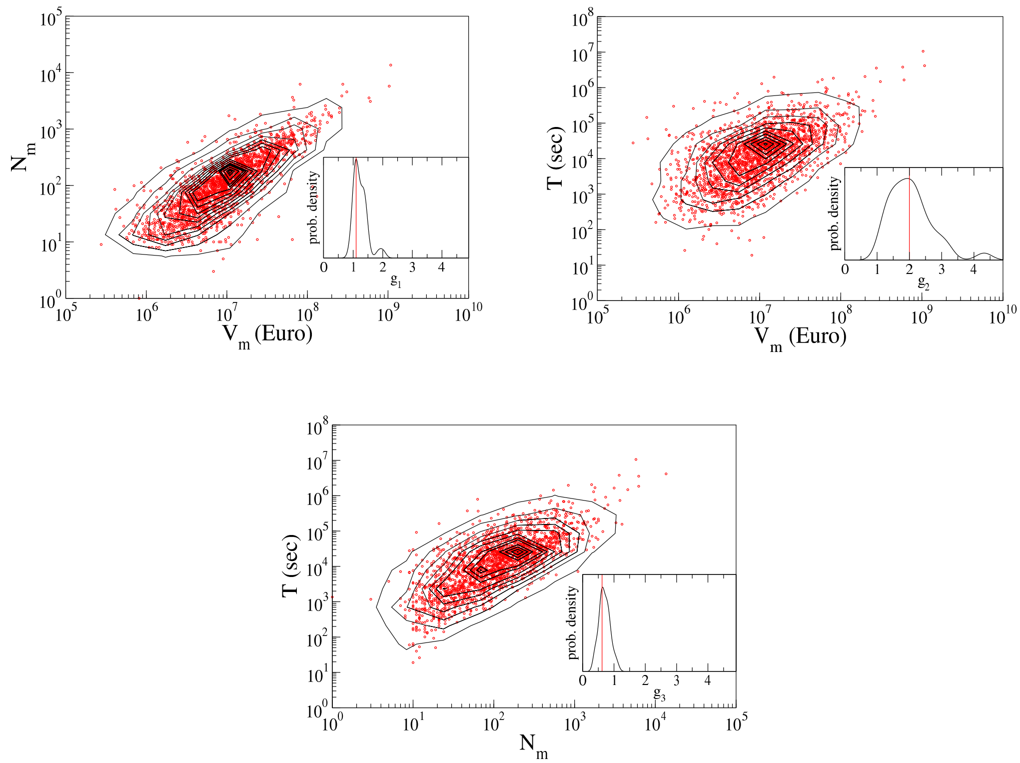

FIG. 3: Scatter plots of the variables T, N_m, and V_m for Telefónica. The black lines are contour lines of the bivariate probability density function. The insets show the probability density functions of the three exponents g_1, g_2, and g_3 describing the allometric relations of Eq. 1 computed on the patches of individual firms with at least 10 patches. The red vertical lines indicates the values of the scaling exponents computed in the pool of all firms and reported in rows 4-6 of Table I. It is worth noting that the dispersion of g_2 is significantly larger than the one for the other two exponents.

dimensional manifold which has one dimension much larger than the other. The fact that the first eigenvalue is large indicates that one factor dominates the trading strategy. The allometric relations of the three variables associated with the first eigenvalue of the PCA provides an estimation of the exponents ($g_1 \simeq 1.2$, $g_2 \simeq 1.8$, and $g_3 \simeq 0.67$ for Telefónica) which, differently than in the bivariate case, are of course coherent among them and only slightly different from the ones obtained from the bivariate analysis.

We now go back to the problem of assessing the role of firm heterogeneity. The first scientific question is: Is the fat tailed distribution of T, N_m, and V_m due to the fact that individual firms place heterogeneously sized packages or is this an effect of the aggregation of many different firms together? To answer this question we test the hypothesis that the patches identified for a given firm trading a given stock are lognormally distributed. The test (see Table I) shows that for most of the trading firms we cannot reject the hypothesis that the patches have characteristics sizes distributed as a lognormal. Since we reject the lognormal hypothesis for the pool obtained by considering all the firms, we conclude that the power law distribution of T, N_m and V_m is due to an heterogeneity in patch scale *between* different firms rather that *within* each firm. The second scientific question about concerns the role of firm heterogeneity for scaling laws. To assess the role of heterogeneity, for each firm we compute the exponents g_1, g_2, and g_3 of the bivariate relations of Eq. 1 (see insets of fig. 3). We observe that the exponents obtained for each firm are distributed around the corresponding value of the exponent obtained for the pool. This result indicates that the bivariate allometric relations are not an effect of the aggregation but are observed, on average, also for individual firms.

In conclusion our comprehensive investigation of packages traded at BME shows that heterogeneity of firms has an essential role for the emergence of power law tails in the investment time horizon, in the number of transactions and in the traded value exchanged by packages. Differently, scaling laws between the variables characterizing each package are essentially the same across different firms with the possible exception of the relation between T and V_m perhaps reflecting different degree of aggressiveness of firms.

I. MATERIALS AND METHODS

Our database of the electronic open market SIBE (Sistema de Interconexión Bursátil Electrónico) allows us to follow each transaction performed by all the firms registered at BME. In 2004 the BME was the eight in the world in market capitalization. We consider the stocks Banco Bilbao Vizcaya Argentaria (BBVA), Banco Santander Central Hispano (SAN), and Telefónica (TEF). We also consider only the most active firms defined by the criterion that each firm made at least $1,000$ trades/year and was active at least 200 days per year. The number of firms is 50 (BBVA), 55 (SAN), and 61 (TEF). These firms are involved in $81-86\%$ of the transactions. The investigated period is 2001-2004. We do not consider other stocks because we have verified that the number of detected patches is too small to perform careful statistical estimation.

The series under study is the series of signed traded value. For each firm and for each stock we construct the series composed by all the trades performed by the firm with a value $+v$ for a buy trade and $-v$ for a sell trade, where v is the value (in Euros) of the traded shares.

The method we use to detect statistically the presence of patches is adapted from Ref. [24] where it was introduced to study patchiness non-stationarity of human heart rate. The algorithm works as follows. One moves a sliding pointer along the signal and computes the mean of the subset of the signal to the left and to the right of the pointer. From these mean values one computes a t statistics and finds the position of the pointer for which the t statistics is maximal. The significance level of this value of t is defined as the probability of obtaining it or a smaller value in a random sequence. One then chooses a threshold (in our case 99%) and the sequence is cut if the significance level is smaller than the threshold. The cut position is the boundary between two consecutive patches. The procedure continues recursively on the left and right subset created by each cut. Before a new cut is accepted one also computes t between the right-hand new segment and its right neighbor and t between the left-hand new segment and its left neighbor and one checks if both values of t are statistically significant according to the selected threshold. The process stops when it is not possible to make new cut with the selected significance.

In the present study, we are mainly interested in directional patches, i.e. patches where the trader consistently buys or sells a large amount of shares. In other words we wish to exclude patches in which the inventory of the firm is diffusing randomly, without a drift. To this end for each patch we compute the total value purchased V_b, the total value sold V_s and the total value $V = V_b + V_s$. We then consider a patch as directional when either $V_b/V > \theta$ (buy patch) or $V_s/V > \theta$ (sell patch). The parameter θ can be varied and in the present study we set it to $\theta = 75\%$. We obtain similar results for different values of θ such as 85% and 95%. Finally in the present paper we consider patches with at least 10 trades.

Acknowledgments Authors acknowledge Sociedad de Bolsas for providing the data and the Integrated Action Italy-Spain "Mesoscopics of a stock market" for financial support. GV, FL, and RNM acknowledge support from MIUR research project "Dinamica di altissima frequenza nei mercati finanziari" and NEST-DYSONET 12911 EU project. EM acknowledges partial support from MEC (Spain) throught grants FIS2004-01001, MOSAICO and a Ramón y Cajal contract and Comunidad de Madrid through grants UC3M-FI-05-077 and SIMUMAT-CM

[1] Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263-291 (1979).
[2] Lux, T. & Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **397**, 498-500 (1999).
[3] Pareto, V. *Cours d'Economie Politique* (Lausanne and Paris, 1897).
[4] Zipf, G.K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
[5] Ijiri, Y. & Simon, H.A. *Skew Distributions and the Sizes of Business Firms* (North-Holland, New York, 1977).
[6] Axtell, R.L. Zipf Distribution of U.S. Firm Sizes. *Science* **293** 1818-1820 (2001).
[7] Pushkin, D. & Aref, H. Bank mergers as scale free coagulation. *Physica A* **336** 571–584 (2004).
[8] Gabaix, X., Gopikrishnan, P., Plerou, V. & Stanley, H.E. Institutional investors and stock market volatility. *Quarterly Journal of Economics* **121** 461-504 (2006).
[9] Calder, W. A. Size, Function, and Life History (Harvard University Press, Cambridge, 1984).
[10] Hasbrouck, J. Measuring the information content of stock trades. *J. Fin.* **46**, 179-207 (1991).
[11] Hausman, J.A., Lo, A.W. & MacKinlay, A.C. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* **31**, 319-379 (1992).
[12] Dufour, A. & Engle, R.F. Time and the price impact of a trade. *J. Fin.* **55**, 2467-2498 (2000).
[13] Plerou, V., Gopikrishnan, P., Gabaix, X. & Stanley, H.E. Quantifying stock-price response to demand fluctuations. *Phys. Rev.* E **66**, 027104 (2002).
[14] Lillo, F., Farmer, J.D. & Mantegna, R.N. Master curve for price-impact function. *Nature* **421** 129-130 (2003).
[15] Bouchaud, J.-P., Gefen, Y., Potters, M. & Wyart, M. Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance* **4**, 176-190 (2004).

[16] Chan L.K.C. & Lakonishok, J. The behavior of stock price around institutional trades. *J. Fin.* **50**, 1147-1174 (1995).
[17] Gallagher D.R. & Looi, A. Trading behaviour and the performance of daily institutional trades. *Accounting and Finance* **46**, 125-147 (2006).
[18] Gabaix, X., Gopikrishnan, P., Plerou, V. & Stanley, H.E. A theory of power-law distributions in financial market fluctuations. *Nature* **423**, 267-270 (2003).
[19] Borland, L. & Bouchaud, J.-P. On a multi-timescale statistical feedback model for volatility fluctuations. Preprint at physics/0507073 (2005).
[20] Lillo, F. Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices. *Eur. Phys. J. B* **55**, 453-459 (2007).
[21] Eisler, Z., Kertesz, J., Lillo, F. & Mantegna, R.N. Diffusive behavior and the modelling of characteristic times in limit order executions. Preprint at physics/0701335 (2007).
[22] Farmer, J.D., Gillemot, L., Lillo, F., Mike, S. & Sen, A. What really causes large price changes? *Quantitative Finance* **4** 383–397 (2004).
[23] Sprent, P. The Mathematics of Size and Shape. *Biometrics* **28** 23–37 (1972).
[24] Bernaola-Galvan, P., Ivanov, P.Ch., Amaral, L.A.N. & Stanley, H.E. Scale invariance in the nonstationarity of physiologic signals. *Phys. Rev. Lett.* **87**, 168105 (2001).